Ferguson, R. J., Prieto-Alhambra, D., Walker, C., Yu, D., Valderas, J. M., Judge, A., Griffiths, J., Jordan, K. P., Peat, G., Glyn-Jones, S., & Silman, A. J. (2019). Validation of hip osteoarthritis diagnosis recording in the UK Clinical Practice Research Datalink. *Pharmacoepidemiology and Drug Safety*, *28*(2), 187-193. https://doi.org/10.1002/pds.4673

Title Page

Validation of hip osteoarthritis diagnosis recording in the UK Clinical Practice Research Datalink

Short title: Validation of Hip Osteoarthritis in CPRD

Authors:
Ferguson RJ[1], Prieto-Alhambra D[1], Walker C[2], Yu D[2], Valderas JM[3], Judge A[1], Griffiths J[4], Jordan KP[2], Peat GM[2], Glyn-Jones S[1], Silman AJ[1]

Author Affiliations:
[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford
[2]Primary Care and Health Sciences Department, Keele University
[3]University of Exeter Medical School
[4]Oxford University Hospital NHS Foundation Trust

Corresponding Author:
Dr Rory Ferguson, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD;
rory.ferguson@ndorms.ox.ac.uk

Key words:
Validation
Hip Osteoarthritis
CPRD

Take home messages or key points:
1. The diagnosis of hip osteoarthritis in primary care is never clear cut
2. CPRD provides an acceptable diagnostic accuracy
3. Given the challenges of diagnosis and the natural history of the disorder, dating the onset is subject to misclassification

Word count: 2984 words

There have been no prior postings or presentations of this work.

Acknowledgements: This research project was funded by an NIHR Research for Patient Benefit grant, with grant number HFNXHF00 HF00.AS. We thank Emma Boyle at CPRD for her support in coordinating the study.

Abstract

Purpose

The diagnosis of hip osteoarthritis is subject to several uncertainties, especially in primary care. The aims of this study were to determine (i) the diagnostic accuracy of coding of hip osteoarthritis by primary care physicians in the UK Clinical Practice Research Datalink (CPRD), (ii) the relative influence of radiographic and clinical parameters on diagnostic accuracy, and (iii) the accuracy of the diagnosis date.

Methods

An extract of all patients aged over 65 years, with a Read code for hip osteoarthritis listed between January 1995 and December 2014, was obtained from CPRD. A random sample was selected of 170 participants. A questionnaire concerning data in medical records on relevant clinical and radiographic criteria used to establish the diagnosis of hip osteoarthritis was distributed to primary care physicians of participants. Using diagnostic criteria, we formulated thresholds for diagnosis based on clinical, radiographic and combined grounds.

Results

119 completed questionnaires were returned (70% response rate). The positive predictive value (PPV) of hip osteoarthritis codes, based on radiological criteria, was 79.8%. The PPV, based on clinical criteria, was 79.0%, with substantial but not complete overlap. Overall 12% of diagnoses were not confirmed. In 42% of cases, there was disparity between date of diagnosis in CPRD and the medical record. Median difference in date was +/-425 days (interquartile range, 18-1448 days).

Conclusions

Despite the difficulties in reaching a diagnosis of hip osteoarthritis in primary care, CPRD Read codes have a sufficiently high PPV for most research uses. However, the accuracy of diagnosis date may not be as reliable.

Background

Osteoarthritis of the hip is one of the most prevalent disabling musculoskeletal problems, affecting an estimated 11% of the population (2.46 million individuals) in England[1]. The only effective treatment for end-stage disease is hip replacement[2]: one of the most commonly performed elective surgical procedures in the elderly. Expenditure in the USA alone on hip replacement was estimated to be $15 billion/year[3].

Patients with hip osteoarthritis experience pain across a wide region in the 'bathing trunk' area, although patterns of pain differ and many also present with stiffness, difficulty in moving and joint swelling[4]. The diagnosis is normally confirmed by imaging, typically based on finding evidence of osteophytes, joint space narrowing and sclerosis on plain radiographic films. Such findings at any grade are frequent in elderly individuals, and the threshold for the presence of significant disease is based on the Kellgren-Lawrence (K-L) grading system[5]. Given the high population prevalence of radiographic abnormalities, the clinical importance of the disorder rests on its impact on patients. Thus, non-radiographic features alone have been considered as sufficient for use by the American College of Rheumatology (ACR) for hospital rheumatologists and by the UK National Institute of Clinical Excellence (NICE) for primary care practitioners [6,7].

Making a diagnosis of hip osteoarthritis in primary care is challenging for a number of reasons. Firstly, the clinical presentation varies[8]. Pain arising from the arthritic hip is difficult to differentiate because pain arising from other structures within the 'bathing trunk' area and because pain may be referred to different locations[9]. Secondly, there is only a weak correlation between symptomatic disease and the radiographic severity [10,11]. Thirdly, there is no single best radiographic definition of what constitutes a 'case. A recent systematic review found studies had interpreted the K-L grading system differently, due to the inexact wording in the original system[12]. There is no consensus on the threshold of joint space width for defining osteoarthritis, with values ranging from <1.5 to <4mm[13–16].

There may be a long lag between first presentation with symptoms and the diagnosis[7]. It is unknown when in the course of a 'patient journey' with hip pain, is the diagnosis of osteoarthritis made in primary care.

There is, thus, a need to interrogate large population primary care databases for hip osteoarthritis, to understand the current and future demands for hip joint replacement. Indeed, electronic health records (EHR) provide an excellent resource because they offer a large cohort size, extensive

demographic and comorbidity information, and data on long-term outcomes. However, the validity of such studies relies on the accuracy of the computerised codes. We are not aware of any study which has investigated the validity of the diagnosis of hip osteoarthritis within primary care records. This is of particular relevance given the substantial issues leading to variation in practice in the diagnosis.

This study therefore set out to examine the validity of the recorded diagnoses of hip osteoarthritis in UK electronic primary care records. Specifically, we aimed to determine (i) the positive predictive value of EHR codes to identify 'true cases' of hip osteoarthritis, (ii) the relative influence of radiographic and clinical parameters on diagnostic recording, and (iii) the potential for error in using electronic recorded date of diagnosis.

Methods

Data set

The Clinical Practice Research Datalink (CPRD) is a UK electronic database of anonymised longitudinal medical records from primary care[17]. The data comprise approximately 14 million patient records of whom around 5.4 million are currently alive and registered at 660 primary care practices throughout the UK[18]. Records are stored via Read codes and contain detailed clinical information on symptoms, diagnoses, investigations, prescriptions and hospital referrals, as well as basic sociodemographic characteristics as entered by the GP. Read codes are a hierarchical clinical coding system of over 80 000 terms that are used in general practice in the UK[19]. Further descriptions of CPRD are published elsewhere[20,21]. The study population consisted of a random sample of the 34656 living individuals registered in CPRD, aged >65 years, with a GP coded diagnosis of hip osteoarthritis made between 1 January 1995 and 31 December 2014, and with a minimum one year of data being available before the date of diagnosis. The Read codes for hip osteoarthritis are shown in Appendix I.

Study design

The accuracy of the hip osteoarthritis diagnosis was assessed in a random sample of the study population by questionnaire survey seeking clinical and radiographic information recorded in the patient's medical records. The questionnaire was mailed by CPRD to the GP of each patient in the sample. This study was approved by the Independent Scientific Advisory Committee for MHRSA database research (ISAC protocol 17_024R).

The questionnaire incorporated items that would allow (i) a radiographic diagnosis and (ii) a clinical diagnosis based broadly on the radiographic and clinical features described in the K-L, ACR and NICE diagnostic schemes. The questionnaire first sought concordance between the CPRD and the GP record, with a specific question on whether the GP patient record had ever had a diagnosis of hip osteoarthritis noted. If so, the questionnaire asked for (i) the date of diagnosis, (ii) whether the diagnosis was confirmed in secondary care, (ii) whether the diagnosis was subsequently revised. The second set of questions focussed on the presence or absence of the specific items used to make the diagnosis. The questions covered symptoms reported by the patient, findings from physical examination, and results from any hip radiographs that were recorded prior to the diagnosis of hip osteoarthritis. The questionnaire is shown in Appendix II. A returned questionnaire was considered invalid if it was returned blank.

Sample size

The study population size was determined by a sample size calculation. To demonstrate sufficient validity, we proposed a PPV of 70% or greater would be required. Based on previous CPRD validation studies for other diagnoses, that obtained additional information from GPs[22], we predicted that 80% of cases would be confirmed as valid. For the lower boundary of the 95% confidence interval of the proportion of valid cases to be above 70%, we calculated that we would need at least 85 returned questionnaires. Based on previous CPRD validation studies, we predicted a questionnaire response rate of 50%, therefore 170 participants were randomly selected. Participants were registered at 138 GP practices (112 practices with 1 participant; 20 practices with 2 participants; 6 practices with 3 participants). The GPs of the participants were mailed the questionnaire. A period of three months for return of questionnaires was permitted before the study was ended.

Data analysis

We calculated response rate as the number of the returned valid questionnaires divided by total number of questionnaires sent. One hypothesis is that the GPs who chose to return a completed questionnaire with clinical details that could be checked against their disease coding might be more confident in their diagnosis. We therefore sought to determine what differences there might be between the patients for whom the GPs responded ('responders'), compared to those patients whose GPs did not respond ('non-responders'), to provide some indication of the background osteoarthritis risk in these two groups of patients. An assessment of both participation bias (comparison of 'responders' versus the whole CPRD cohort) and of non-response bias (comparison of 'responders' versus 'non-responders') was done by comparing age, gender, body mass index (BMI), socio-economic

status (SES) and comorbidities. SES was derived from the Index of Multiple Deprivation quintile of the GP practice region[23]. Comorbidities were scored with the Charlson Comorbidity Index[24].

An algorithm flow chart (Figure I) was used to make the diagnostic assignment of hip osteoarthritis. Cases were defined as radiologically confirmed cases of hip osteoarthritis if the report of radiographs reported evidence of 'osteoarthritis', 'osteophytes' or 'joint space narrowing'. Cases were defined as clinically confirmed cases if patients reported pain, stiffness or restriction of movement, or if physical examination found a reduced range of hip flexion or internal rotation. Cases were defined as unconfirmed cases if these radiological or clinical parameters were not met.

Patients could be classified as satisfying (i) radiographic criteria, (ii) clinical criteria, (iii) *either* radiographic *or* clinical criteria, and (iv) *both* radiographic *and* clinical criteria. The PPV refers to the proportion of the cases that were confirmed as actual cases, based on the information provided by the GP in the questionnaire. We calculated PPV as the number of confirmed cases divided by the total number of cases for whom a valid questionnaire was received from the patient's GP. Binomial 95% confidence intervals were calculated for the PPV. In patients with a difference in the date of diagnosis recorded between CPRD and the GP patient record, the number of days between the two dates was calculated. The distribution of the difference in date in those without an exact date match is displayed graphically.

Results

A total of 125 (73.5%) questionnaires were returned. Six of the returned questionnaires were blank and therefore invalid, because GPs reported they could not access required data (in 5 cases GPs had changed computer software; in 1 case the patient was no longer registered at the practice), yielding a final response rate of 70.0% (119/170). The response rate was higher amongst patients with a more recent diagnosis (date earlier/later than the median diagnosis date): 74% vs 66%.

The comparison of demographic characteristics between the three populations of (i) the entire CPRD cohort with hip osteoarthritis, (ii) responders, and (iii) non-responders to the 170 questionnaires are shown in Table I. The characteristics were broadly similar. The non-responders were more likely to be male, but had very similar mean age, BMI and comorbidity status. Indeed, the responders had a socio-economic profile very close to the entire CPRD cohort with hip osteoarthritis.

In all, 95 of the 119 cases were confirmed based on radiological parameters, yielding a PPV of 79.8% (95% confidence interval (CI): 72.6%-87.0%), whilst 94 cases were confirmed based on clinical parameters, yielding a PPV of 79.0% (95% CI: 71.7%-86.3%). The level of agreement between radiological and clinical confirmed cases is shown in Table II, with a Kappa statistic of 0.46, p value <0.001. If either clinical or radiographic criteria were required to confirm diagnosis, the PPV was 88.2% (95% CI, 82.4%-94.0%). If both clinical and radiographic criteria were required to confirm diagnosis, the PPV fell to 70.5% (95% CI, 62.3%-78.7%).

Of the 14 unconfirmed cases, in 12 cases there was no record of a diagnosis of osteoarthritis ever having been made. One case was unconfirmed because the diagnosis was subsequently changed to trochanteric bursitis. One case was unconfirmed because radiological and clinical criteria were not met; however, the GP reported that the diagnosis of hip osteoarthritis was subsequently confirmed in secondary care.

In 104 of the total 105 confirmed cases (99.0%) the date of diagnosis was recorded in the GP patient record. In the majority of cases (60/104), the date of diagnosis was identical to that on CPRD. However, in the other 44 cases (42.3%) there were discrepancies in diagnosis date in both directions; in the majority (32/44), the GP patient record recorded an earlier date of diagnosis than the CPRD record. For many patients, there were substantial differences between the two dates (Figure II). Median difference in diagnosis date was +/-425 days (interquartile range, 18-1448 days). In only 11 (26%) of patients was the difference less than 30 days; in others (24 cases) the difference extended to over 1 year, with 16 cases of earlier diagnosis in the GP patient record and 8 cases of earlier diagnosis in CPRD.

Discussion

The results of this validation study show that the diagnosis of hip osteoarthritis recorded in the CPRD has high PPV when assessed by radiological or clinical criteria (79.8% and 79.0% respectively). These findings suggest that, using the criteria described here, the diagnosis of hip osteoarthritis in patients aged over 65 years made by GPs and recorded in the CPRD is accurate. The level of accuracy of the diagnosis varied depending on the standard chosen; however, relying on radiological evidence or clinical evidence alone produced very similar rates. By contrast caution should be taken in using the diagnosis date for research because the accuracy is mixed.

The presence of 12 CPRD hip osteoarthritis cases having no recorded diagnosis in the GP patient record was unexpected. Possible explanations are that there was an error in completing the questionnaire by the GP, that the code was entered into CPRD in error, or that after diagnosis was made and recorded in CPRD, the GP subsequently edited the patient record. Of note, one of the cases unconfirmed by radiological or clinical criteria but was reported to be confirmed in secondary care. The lack of evidence in the GP patient record to confirm the diagnosis may be because information was not passed back from secondary care to primary care. Our analysis plan was to confirm cases based on defined radiological or clinical criteria, so this patient was treated in the analysis as an unconfirmed case because these criteria were not met. Additionally, not all patients are seen in secondary care, so permitting the confirmation of diagnoses from secondary care would have introduced bias into our results.

To our knowledge, no studies have previously investigated the PPV of hip osteoarthritis diagnosis in the CPRD dataset. We identified one study that investigated the validity of osteoarthritis diagnoses recorded in health records in the British Colombia (Canada) Ministry of Health administrative database[25]. They found that in patients diagnosed with knee osteoarthritis by one physician, the PPV of a valid diagnosis of knee, hip or hand osteoarthritis based on clinical criteria was 82% (95% CI, 71-89%). This is consistent with our findings.

Selecting appropriate criteria to determine if a diagnosis of hip osteoarthritis is valid or not was fundamental. Our approach to validate the diagnosis was based on the ways that the diagnosis of hip osteoarthritis is made in practice based on using both radiological and clinical information.

Our approach to radiologically confirming the diagnosis allowed the presence of mention of 'osteoarthritis' or of osteophytes or joint space narrowing as sufficient. The latter are the radiographic features required for K-L grading to radiologically define osteoarthritis. We were unable to obtain the radiographs to formally K-L grade them and it was unlikely such grading would be available on routine GP records. Our radiologically confirmed cases did not also require any further clinical evidence. We believe this approach is justified because x-rays of the hip would most likely have been performed to investigate symptoms in the hip region. However, it is possible that in some cases x-rays were performed for another clinical reason (for example, to investigate general surgical pathology), and that radiographic evidence of hip osteoarthritis was an incidental finding, leading to a diagnosis in the absence of clinical disease.

Our approach to clinically confirming the diagnosis of hip osteoarthritis, allowed the presence of mention of either pain, stiffness or restriction of movement, or reduced range of hip flexion or internal rotation as sufficient. This differs from the NICE clinical diagnostic criteria. NICE require activity-related joint pain and less than 30 minutes of morning stiffness in patients 45 years or over. In this retrospective study, we took a pragmatic view, after discussing with primary care colleagues, that the quality of the data in routine records would not be of sufficient quality to allow robust assignment using criteria such as those of NICE or ACR. It was unlikely that there would be any accurate measure of, for example, range of hip flexion and internal rotation, these being good markers of hip osteoarthritis[26].

Thus, it is acknowledged that the basis for 'proving' the diagnosis of hip osteoarthritis from retrospective primary care records is limited by the available data. In that context, it is considered that the PPV from radiographically confirmed cases is more valid.

There was potential for bias between the quality of coding of GPs that responded and those that did not. We have demonstrated that there were no significant differences between the patients based on age, gender, comorbidity or their GP practice index of multiple deprivation. Of note, the profile of social demographic status of responders closely mirrored the entire CPRD cohort. However, we are unable to exclude the possibility that in some of the cases of unreturned questionnaires, GPs may have chosen to not respond on finding a lack of evidence to confirm the diagnosis of hip osteoarthritis.

The sample size was sufficient to demonstrate that the diagnosis was valid using radiological and clinical parameters separately; however, a larger sample size would have enabled a more accurate PPV calculation, with a smaller 95% CI. As with many validation studies, this study was limited by costs. Costs were £35 for each questionnaire distributed by CPRD and £55 for each questionnaire returned by GPs. The period for responses was limited to three months due to time constraints and because it was predicted that the majority of responses would be received within two months. This transpired: 84% (105/125) returned questionnaires were returned within two months.

We attempted to ensure that only incident cases of hip osteoarthritis during the study period were included. Our approach to achieve this was to exclude patients for whom there was a mention of hip osteoarthritis in the 12 months prior to diagnosis in the CPRD system. However, we cannot exclude the possibility that some of the patients in this analysis may have had a diagnosis of hip osteoarthritis previously. This could occur in two ways. Firstly, in the UK patients are free to move between different

primary care providers. The CPRD database can only capture clinical events during the time when an individual patient is registered with a CPRD practice, and therefore some patients may have had a diagnosis of hip osteoarthritis previously in a non-CPRD practice. Secondly, patients may have had a diagnosis of hip osteoarthritis in a CPRD practice more than one year previously and had not consulted their GP for at least a year. On investigation of the date of diagnosis, we found 16 cases where the date of diagnosis in the GP patient was more than one year earlier than in the CPRD record. These 16 cases may represent cases where the date of diagnosis in CPRD was not accurate or prevalent cases where an earlier diagnosis was recorded in CPRD but was missed because we only investigated one year prior to the diagnosis within our study period. Because of this uncertainty, we elected not to exclude these 16 cases. We believe it is unlikely that either scenario is likely to be of material impact to the results of the coding accuracy analysis. However, were it possible to identify and exclude such patients, the disparity in date of diagnosis in CPRD and the GP patient record would likely be lower."

This study aimed to assess the accuracy of "new" diagnoses of hip osteoarthritis using the Read Codes provided by primary care practitioners on the CPRD database. Although there are estimates of the point prevalence of hip osteoarthritis from population surveys, it was not possible for us to derive any comparative estimate of the prevalence of OA in the CPRD population, because our patient cohort only constituted new diagnoses in primary care. Further studies may attempt to investigate the proportion of hip OA cases seen in UK primary care that are not diagnosed or coded by the GP, although this may prove challenging because there are no "gold-standard" data to compare primary care data with. Indeed, there are studies from many disorders in the literature where CPRD itself has been used as such a gold-standard.

A key limitation is that from our data we can calculate PPV of the diagnosis in CPRD, but not sensitivity. There may exist many cases of patients in the CPRD that have hip osteoarthritis but do not have it diagnosed or recorded. It was not possible to identify these patients, thus a sensitivity calculation was not possible. Future studies on hip osteoarthritis in the CPRD should consider this potential source of bias.

It is also important to consider whether there were codes used by some primary care practitioners for registering cases of hip osteoarthritis that were not considered in this analysis. For example, we did not include those codes that covered "osteoarthritis of the pelvic area" but had not mentioned hip. It would be important in future studies to determine if such codes are widely used by some primary care practitioners for hip osteoarthritis. However, whilst our exclusion of considering such cases would

under-estimate the total number of hip OA patients, a plausible conclusion is that the levels of clinical and radiographic confirmation of such cases would not necessarily be different from those where the primary care practitioner preferred to use the term hip. This however remains to be proven.

The cohort of patients was restricted to patients aged over 65 years. This is because there is greater diagnostic uncertainty in patients aged under 65, and therefore more complex to validate. The findings of this paper should not be assumed to apply to diagnoses of hip osteoarthritis made in younger patients.

Conclusion

In conclusion, the recorded diagnosis of hip osteoarthritis was confirmed in a high proportion of patients aged over 65 years within CPRD, and therefore Read codes for hip osteoarthritis amongst all patients aged over 65 in the CPRD dataset are likely to accurately identify true cases. However, accurate diagnosis timing remains difficult to determine.

Table I: Demographic characteristics of patients aged >65 years with hip osteoarthritis recorded within CPRD, and responders and non-responders to the GP validation questionnaire

	Entire cohort		Responders		Non-responders	
	Number	%	Number	%	Number	%
Cohort size	34, 656		119		51	
Gender						
Male	14183	40.9	55	46.2	26	51.0
Female	20473	59.1	64	53.8	25	49.0
Age (mean, SD)	75.2	11.5	73.7	11.7	74.3	11.4
BMI (mean, SD)	28.4	5.3	28.6	5.1	28.5	6.2
CCI (mean, SD)	0.36	0.95	0.40	0.93	0.35	1.05
IMD						
Quintile 1: Least deprived	5688	26.4	17	27.4	5	20.8
Quintile 2	5342	24.8	14	22.6	7	29.2
Quintile 3	4664	21.6	12	19.4	1	4.2
Quintile 4	3756	17.4	13	21.0	8	33.3
Quintile 5: Most deprived	2111	9.8	6	9.7	3	12.5
Missing	13095	-	57	-	27	-

BMI: body mass index (kg/m^2); CCI: Charlson Comorbidity Index; IMD: Practice level Index of Multiple Deprivation, 2004

Table II: Overlap of cases confirmed by radiological and clinical parameters

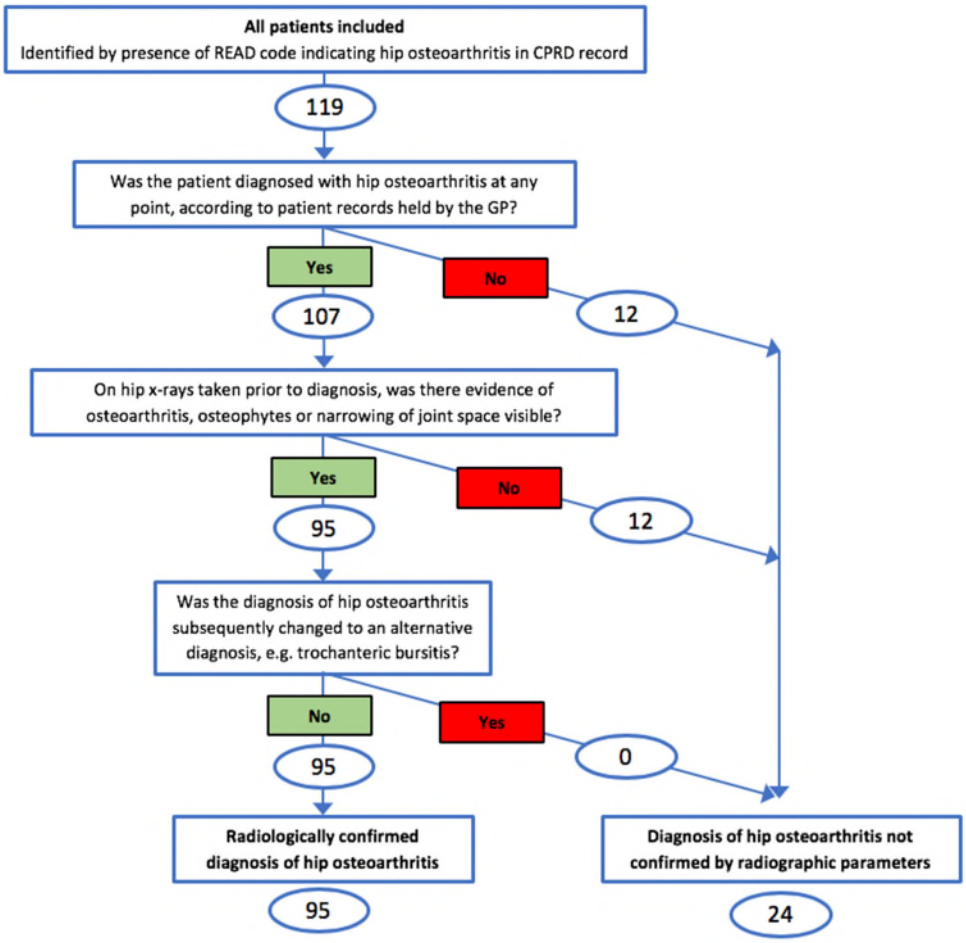
		Clinical		
		Confirmed	Unconfirmed	All
Radiological	Confirmed	84 (70.5%)	11 (9.2%)	95 (79.8%)
	Unconfirmed	10 (8.4%)	14* (11.8%)	24 (20.2%)
	All	94 (79.0%)	25 (21.0%)	119 (100%)

Kappa statistic = 0.46, p value <0.001

*One case was unconfirmed by radiological and clinical parameters; however, the GP reported that the case was confirmed in secondary care.

Figure I: Algorithms for diagnosis based on radiologic and clinical parameters, based on returned questionnaire answers

(a) Algorithm for diagnosis based on radiologic parameters



(b) Algorithm for diagnosis based on clinical parameters

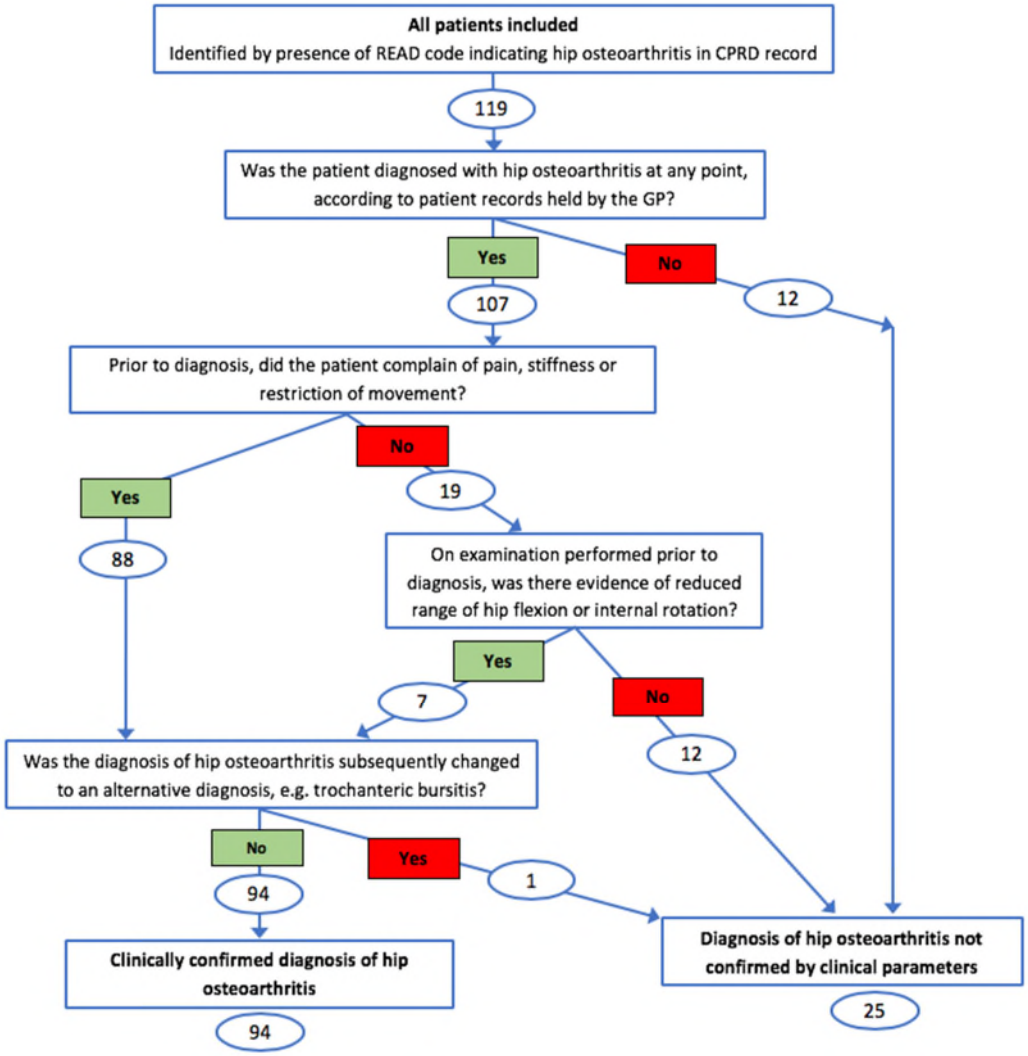
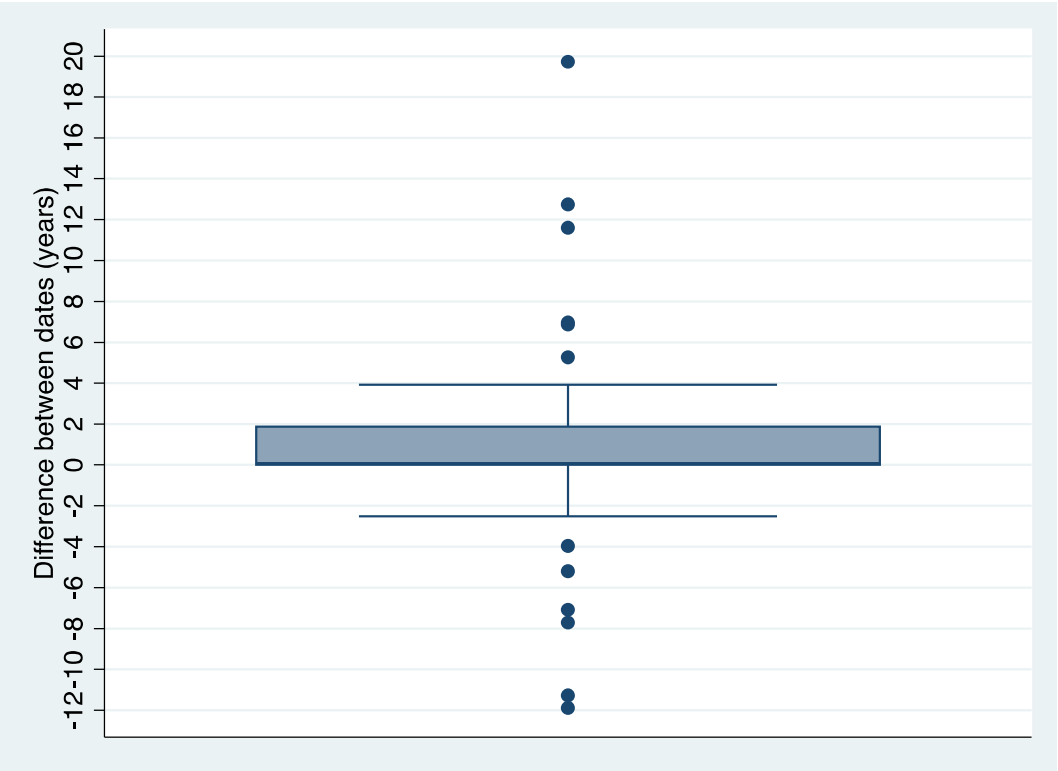


Figure II: Box and whisker plot of the time in years between dates, of patients with a different date of diagnosis between CPRD and GP patient record. (N=44).



References:

1 Arthritis Research UK and Imperial College London. The musculoskeletal calculator. www.arthritisresearchuk.org/mskcalculator. 2014.

2 Pivec R, Johnson AJ, Mears SC, Mont MA. Hip arthroplasty. *Lancet* 2012; **380**: 1768–77.

3 Lavernia CJ, Hernandez VH, Rossi MD. Payment analysis of total hip replacement. *Curr Opin Orthop* 2007; **18**: 23–7.

4 Arthritis Health. Hip Osteoarthritis Symptoms and Signs. https://www.arthritis-health.com/types/osteoarthritis/hip-osteoarthritis-symptoms-and-signs (accessed Dec 18, 2017).

5 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957; **16**: 494–502.

6 No Authors Listed. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. *Arthritis Rheum* 2000; **43**: 1905–15.

7 National Institute for Health and Clinical Excellence. Osteoarthritis: Care and Management Guidelines. 2014. https://www.nice.org.uk/guidance/cg177 (accessed Aug 11, 2017).

8 Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. *Clin Orthop Relat Res* 2016; **474**: 1886–93.

9 Birrell F, Lunt M, Macfarlane GJ, Silman AJ. Defining hip pain for population studies. *Ann Rheum Dis* 2005; **64**: 95–8.

10 Kinds MB, Welsing PMJ, Vignon EP, *et al.* A systematic review of the association between radiographic and clinical osteoarthritis of hip and knee. *Osteoarthr Cartil* 2011; **19**: 768–78.

11 Kim C, Nevitt MC, Niu J, *et al.* Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. *BMJ* 2015; **351**: h5983.

12 Schiphof D, Boers M, Bierma-Zeinstra SMA. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. *Ann Rheum Dis* 2008; **67**: 1034–6.

13 Roach KE, Persky V, Miles T, Budiman-Mak E. Biomechanical aspects of occupation and osteoarthritis of the hip: a case-control study. *J Rheumatol* 1994; **21**: 2334–40.

14 Lindberg H, Danielsson LG. The relation between labor and coxarthrosis. *Clin Orthop Relat Res* 1984; : 159–61.

15 Croft P, Coggon D, Cruddas M, Cooper C. Osteoarthritis of the hip: an occupational disease in farmers. *BMJ* 1992; **304**: 1269–72.

16 Thelin A, Jansson B, Jacobsson B, Ström H. Coxarthrosis and farm work: a case-referent study. *Am J Ind Med* 1997; **32**: 497–501.

17 Clinical Practice Research Datalink. https://www.cprd.com/intro.asp (accessed Sept 6, 2017).

18 Quint JK, Mullerova H, DiSantostefano RL, *et al.* Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open* 2014; **4**: e005540–e005540.

19 Chisholm J. The Read clinical classification. *BMJ* 1990; **300**: 1092.

20 Wood L, Coulson R. Revitalizing the General Practice Research Database: plans, challenges, and opportunities. *Pharmacoepidemiol Drug Saf* 2001; **10**: 379–83.

21 Wood L, Martinez C. The general practice research database: role in pharmacovigilance. *Drug Saf* 2004; **27**: 871–81.

22 Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69**: 4–14.

23 UK Government. Indices of Deprivation. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (accessed Dec 22, 2017).

24 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40**: 373–83.

25 Rahman MM, Kopec JA, Goldsmith CH, Anis AH, Cibere J. Validation of Administrative Osteoarthritis Diagnosis Using a Clinical and Radiological Population-Based Cohort. *Int J Rheumatol* 2016; **2016**: 1–7.

26 Birrell F, Croft P, Cooper C, *et al.* Predicting radiographic hip osteoarthritis from range of movement. *Rheumatology (Oxford)* 2001; **40**: 506–12.

Supplementary material

Appendix 1: Read codes for hip osteoarthritis

1. N053512 Hip osteoarthritis NOS
2. N05z511 Hip osteoarthritis NOS
3. N05zJ00 Osteoarthritis NOS, of hip
4. Nyu2100 [X]Other primary coxarthrosis
5. N051900 Primary coxarthrosis, bilateral

Appendix 2: Questionnaire

Practice ID:	Vision Patient ID:	Sex:	Birth Year:

Questionnaire into Validation of Hip Osteoarthritis

1. Diagnosis

a) Was the patient diagnosed with osteoarthritis of the hip?
□ Yes
If Yes, please provide the diagnosis date: __ / __ / ____
□ No
If No, no further responses to the questionnaire are required

b) Was the diagnosis confirmed in secondary care?
□ Yes
□ No

c) Was the diagnosis subsequently changed to an alternative diagnosis?
□ Yes
□ No

2. Presentation

a) Prior to the diagnosis of osteoarthritis of the hip, did the patient attend with any hip related symptoms?
□ Yes
□ No
□ Unknown *(If No or Unknown, please proceed to question 3)*

b) Please indicate which symptoms the patient complained of *(Please tick all that apply)*:
□ Pain
□ Stiffness
□ Restriction of movement
□ Other (please indicate): _____

3. Physical Examination

According to the patient's electronic healthcare record (including hospital letters), please indicate which best apply based on the last joint examination made prior to diagnosis or hip osteoarthritis *(Please tick all that apply)*:

□ Normal range of joint movement
□ Reduced range of hip flexion
□ Reduced range of hip internal rotation
□ No record of joint examination

4. Imaging Investigations

According to the patient's electronic healthcare record (including hospital letters), please indicate which apply based on hip or pelvis X-rays/scans taken prior to diagnosis of hip osteoarthritis *(Please tick all that apply)*:

□ Osteoarthritis
□ Osteophytes
□ Narrowing of hip joint space

Many thanks for your time in completing this questionnaire. Please return it as soon as possible to the CPRD in the freepost envelope provided.